Long-term Effects of Parental Migration on Income: Evidence from Indonesia

Author: Max Bahar

Advisor: Prof. Paul Cichello

Senior Honors Thesis

Department of Economics

Boston College

May 2021

I Abstract

Migration is becoming increasingly common in the developing world. A growing body of literature seeks to address the effects of migration on the families of migrants; namely, the effects of migration on the children of migrants. This study uses the Indonesia Family Life Survey (IFLS) panel dataset to quantify the long-term impacts of experiencing parental migration as a child (aged 5 to 18) on the income of working Indonesian adults. To address the issue of endogenous migration, the out migration rate of an individual's birth Kabupaten (Regency) is used as an instrumental variable. The results of this study indicate that the proposed instrumental variable strongly predicts an individual's parental migration status. However, the wide standard errors on the coefficients of interests prohibit any conclusive remarks to be made on the effects of parental migration on future income. This study illustrates how extensive panel datasets, such as the IFLS, can facilitate analyses on the long-term effects of parental migration. The author recommends further research on the topic of parental migration be performed on other outcome variables such as education, measures of health, and subjective wellbeing.

II Acknowledgements

First and foremost, I would like to thank my thesis advisor Professor Paul Cichello. His guidance and knowledge were instrumental to the completion of this thesis. I would also like to extend my thanks to Professor Robert Murphy and the Economics Department at Boston College for fueling my interest in economic research. Finally, I would like to personally thank my friends and family for showing their continued support and excitement for my efforts at Boston College. Namely, my mother, Taty The; my siblings, Myles, Carleen, Marvyn, and Martin; my girlfriend, Victoria Pusung; and my friends, Sireesh Vinnakota, Tanya Walia, Sathvika Dabbi, Christina Be, Bao Dinh, Ziyu Fan, and many more who have helped shape me throughout my four years at Boston College.

1 Introduction

In recent years, many developing countries have experienced increased levels of migration, both internal and international. Due to its prevalence, numerous studies have tried to quantify the effects of migration on the locations that send and receive migrants (Wong and Chang, 2010; Nathan, 2011; Brücker and Jahn, 2011). However, there is a burgeoning body of literature that seeks to study the effects of migration on the outcomes of other members in a family; namely, the effects of parental migration on children's outcomes.

It is widely agreed upon that the net effect of parental migration on a child's outcomes is ambiguous (Antman, 2012a; Demurger, 2015). Remittances from a migrating parent may improve children's outcomes. But, the absence of a parent or culture shock caused by migrating may be detrimental to the child. Although the literature still remains divided on the net effect of parental migration, many studies agree that these effects are influenced by the gender of both the migrant and the child. Additionally, researchers must tackle the issue of endogeneity due to migrants self-selecting into migration. This endogenous variable issue is commonly solved with propensity score matching, fixed effects, and/or instrumental variables.

Most of the existing literature on the effects of parental migration on children's outcomes examine parental migration in the context of labor migration where the child is left behind. Furthermore, studies are limited to observing short-term measures of children's outcomes, usually in terms of education or health. This paper seeks to extend the current literature in three key ways. First, this paper will attempt to quantify the long-term effects of parental migration on children by linking individuals' parents' migration histories obtained in 1993 – 2000 with individuals' income data recorded in 2014. Second, this paper will expand the types of outcomes studied by using log of real annual income as the main dependent variable. Third, this paper broadens the definition of migration to capture the aggregate effects of all types of parental migration on children's outcomes.

This study uses the Indonesia Family Life Survey (IFLS) to quantify the long-term effects of parental migration on children's log of real annual income. The IFLS is an extensive panel dataset fielded in five waves in 1993, 1997, 2000, 2007, and 2014. To tackle the issue of endogenous migration, I use the out migration rate of an individual's birthplace Kabupaten[1] as an instrumental variable.

I find inconclusive results regarding the causal effects of parental migration on an individual's future income due to the wide standard errors on the coefficients of interest. This may be attributed to the relatively weak explanatory power of the regression model and measurement error commonly present in income data. Consistent with existing literature, I find that the out migration rate at the Kabupaten level strongly predicts whether or not an individual has experienced parental migration, indicating that it is a strong instrumental variable. Furthermore, I find that individuals with better educated parents and a smaller number of household members are more likely to experience parental migration. Contrasting existing literature, I find that parental migration mostly consists of non-labor migration (94.6%) and migration where the child is brought along (76.2%).

The main benefit of this paper is to illustrate that it is possible to quantify the long-term effects of migration using large-scale longitudinal datasets such as the IFLS. However, future studies may choose to study other outcomes that suffer from less noise, such as educational attainment, measures of health, and subjective wellbeing. Additionally, future research focusing on Indonesian parental migration should pay closer attention to non-labor migration and perform further investigation into the specific motivations in this category of migration.

[1] A Kabupaten or Regency is a second-level administrative division of Indonesia, directly under a province. Comparing to the US, Kabupatens are similar to counties, while provinces are analogous to states.

2 Literature Review

2.1 Indonesian Migration

Indonesia has experienced an influx of internal and international migration in recent years. According to UNESCO, almost 9.8 million individuals were estimated to be temporary internal migrants between 2005 and 2010, while the number of international migrants is estimated to be around 4.5 million. Trends from 2015 suggest that most internal migrants are in the 15 to 34 age bracket, better educated, and motivated by labor-related reasons (UNESCO, 2018).

These observations are backed up by a study examining internal migration in Indonesia conducted by Pardede, McCann, and Venhorst. Using a logit model and the longitudinal Indonesia Family Life Survey dataset, they find that individuals are more likely to have migrated if they are in the 15 to 24 age bracket and have attained a higher level of education. This study adds that individuals are more likely to migrate if they have never been married, live in a household with only 1 to 4 people, and have no dependents (Pardede, McCann, and Venhorst, 2020).

Although existing evidence implies that most migrants tend to be childless, the sheer number of individuals involved in migration in Indonesia suggest that a large number of families in Indonesia experience parental migration. Therefore, it is important to uncover the aggregate effects of a parent's migration on a family's wellbeing. However, existing literature reveals mixed results.

2.2 Effects of Parental Migration

Most studies quantifying the effects of parental migration do so by examining a child's short-term educational and health-related outcomes. It is widely acknowledged that the net effect of parental migration on a child's short-term educational and health-related outcomes is ambiguous. For example, it is possible that a child experiencing labor-related parental migration will receive more educational and health-related inputs, such as tutoring or a more nutrition-dense diet, due to remittances sent by the parent(s). This example illustrates how parental migration improves a child's short-term educational and health-related outcomes.

However, it may also be the case that a child who is brought along with the migrating parent experiences psychological stress from moving which worsens their performance at school and becomes detrimental to their mental health. This example illustrates how parental migration worsens a child's short-term outcomes.

The studies collected in the next sections are compiled from a variety of different countries and contexts, ranging from international parental migration in the Philippines (Yang, 2008) to rural-urban migration of Chinese villagers (Bai et al, 2017). Although the circumstances of migration in each situation are unique, these studies all shed light on the ways parental migration can affect children. Before diving into the literature, we must first tackle the main econometric challenge in migratory studies: the endogeneity of migration.

2.3 Endogeneity of Parental Migration

Researchers looking to study the effects of parental migration on children's outcomes must be wary of the endogeneity of migration. The decision for individuals to participate in migration is not random. In other words, there is a selection bias problem. Furthermore, unobservable variables that influence migratory decisions are likely to be correlated with the outcomes of children. This introduces an omitted variable problem (Social Science Research Council, 2009; Antman, 2012a). One common way to tackle the issue of endogeneity is to use an instrumental variable approach.

Studies focusing on migration have employed a wide array of instrumental variables, including historic migration rates (Hanson and Woodruff, 2003; McKenzie and Rapoport, 2013), weather changes in the origin (Meng and Yamauchi, 2017), variables correlated to the demand of migrant workers (Cortes, 2015), and the share of children with migratory parents studying in the same school district (Zhao et al., 2014; Mao, Zang, and Zhang, 2020). However, as is the difficulty with most instrumental variables, one cannot easily prove that the instrument is uncorrelated with the outcome variables of interest.

2.4 Positive Effects

One way that parental migration can positively affect a child's short-term educational and health-related outcomes is through an income effect. Remittances from the migrating parent will relax the household budget, allowing children to receive more investments in their education and health. Additionally, demand for child labor in the household may decrease as a result of remittances, allowing the child to invest more of their time in their education and avoid additional stressors to their health (Antman, 2012a; Demurger, 2015). Several studies find positive net effects of parental migration on children's outcomes.

Bai et al. uses a two-wave panel dataset containing more than 13,000 students in 130 primary schools in rural Northwestern China to find that parental migration positively impacts the academic performance of left-behind children, measured through changes in standardized English test scores. They use a difference-in-difference approach and propensity score matching to account for the endogeneity of parental migration. However, they are unable to prove that this positive effect is caused by increases in income (Bai et al, 2017).

On the other hand, Yang uses data from overseas labor migrants from the Philippines and exchange rate shocks in the aftermath of the Asian financial crisis to discover that migrant income shocks significantly affects the investment decisions of households at the origin, particularly educational investments. The study finds that households that experience favorable exchange rate shocks raise expenditure on schooling, keep children in school longer, and take children out of the labor force. These results suggest that migrant remittances have a positive effect on children back home (Yang, 2008).

2.5 Negative Effects

However, parental migration can also be detrimental to children's short-term educational and health related outcomes. For example, the absence of a parent in a household can cause disruptions in family life, which can be detrimental to children's outcomes. The disappearance of a caregiver and/or voice of authority may cause the child to perform worse at school due to increased emotional turmoil at home and additional household responsibilities and chores. Children of migrant parents may also develop psychological issues and unhealthy habits as a result of decreased bonding time with their parents and a lack of hands-on parental guidance (Antman, 2012a; Demurger, 2015; Zhao et al., 2018). Several studies find that the negative effects of parental migration dominate over its positive effects.

Mao, Zang, and Zhang uses two nationally representative Chinese datasets and an instrumental variable approach to determine the effect of parental migration on both short-term schooling outcomes and long-term educational achievement. After using the share of left behind children within the same school as an instrument, they find that, compared to children of non-migrant parents, left-behind children receive lower cognitive test scores, lower academic scores, and are less likely to go to college (Mao, Zang, and Zhang, 2020). This is one of the few instances that the existing literature tackles the long-term effects of parental migration, measured by the probability of going to college.

Using data from the Young Lives Project for children in Ethiopia, India, Peru, and Vietnam, Nguyen finds that although parental migration increases household per capita expenditure, this does not translate to improved educational or health outcomes for the children. Accounting for the endogeneity of migration by using child fixed-effects, they find that parental migration significantly decreases measures of children's health in India, Peru, and Vietnam, while the effects in Ethiopia are insignificant. Additionally, they find that parental migration decreases cognitive ability test scores in India and Vietnam. (Nguyen, 2016).

2.6 Gendered Effects of Parental Migration

Much of the literature finds that the effects of parental migration differ based on the gender of both the parent and the child. The gendered effects of parental migration can be attributed to differences in household members' preferences over goods and shifts in bargaining power among household members (Antman, 2012a; Demurger, 2015). Existing literature generally finds that, compared to paternal migration, maternal migration is more detrimental to children's educational outcomes. Additionally, paternal migration is found to be more beneficial for girls compared to boys.

In their research paper, Cortes examines the gendered effects of Philippine transnational parental migration on left-behind children's educational outcomes. To study the importance of gender, they compare children whose mothers migrated with children whose fathers migrated. Using shocks to the demand of female migrants as an instrument to predict the migrant's gender, Cortes finds that children of migrant mothers are significantly more likely to lag behind in school compared to children of migrant fathers. Interestingly, additional analysis seems to indicate that this gender gap is driven by parental absence instead of lower remittances (Cortes, 2015).

Also examining educational outcomes, Antman uses data from Mexico to find that paternal Mexico-to-US migration only increases educational attainment for daughters. They estimate that having the father migrate before the child turns 20 increases educational attainment by about 0.73 years for girls. However, they did not find any statistically significant effects for paternal Mexico-to-US migration on boys or any effects of paternal domestic migration on both boys and girls educational outcomes. (Antman, 2012b).

Consistent with the findings of Antman, Hanson and Woodruff find that paternal Mexico-to-US labor migration in rural Mexico only increases the educational attainment for young girls. Namely, for 10-15 year old girls whose mother has less than three years of education, paternal migration is associated with an increase of 0.73 to 0.89 years of schooling (Hanson and Woodruff, 2003).

2.7 Contributions

Based on my literature review, I have noticed that most studies focus on the educational and/or health-related outcomes of children with migrant parents. Often times, measures of these outcomes are taken at the short-term scale. Furthermore, most of the literature restricts parental migration to situations where the child is left behind at the source location and the parent is driven to migrate by labor reasons. I hope to extend the existing literature in the following key ways: First, I will be examining the long-term effects of parental migration on children. The comprehensive Indonesia Family Life Survey dataset will allow me to examine how experiencing parental migration as a child will affect that person as an adult (a gap of more than 20 years). Second, I will expand the types of outcomes studied by observing how experiencing parental migration as a child will affect real annual income as an adult. Third, I will use a more general definition of migration to capture the aggregate effect of all types of parental migration on children's outcomes.

3 Methodology

3.1 Data

Figure 1 – Map of IFLS provinces in Indonesia (RAND, 2020).

To examine the effects of parental migration on long-term real annual income, I will be using data from the Indonesia Family Life Survey (IFLS). The IFLS is an extensive panel dataset following around 30,000 individuals over 5 waves of surveys fielded in 1993, 1997, 2000, 2007, and 2014. Although the data only captures 13 out of the 27 Indonesian provinces, with most concentrated in the island of Java, it is representative of 83% of the national population based on the 1993 Indonesian Census. The IFLS stratifies on provinces and has a total of 321 enumeration areas at the village level (RAND, 2020).

For my study, I will be using waves 1, 2, and 3, fielded in 1993, 1997, and 2000, to obtain migration histories of the individual's parents. Then, I will use the IFLS5, fielded in 2014, to obtain data on the individual's income. Using multiple waves of the IFLS dataset will allow me to measure the long-term effect of parental migration on future income with a gap of 10 or more years since the occurrence of parental migration.

The real annual income of an individual in 2014 (*log real)* is the main dependent variable examined. It consists of formal sector annual wages and bonuses as well as informal sector farm and non-farm enterprise profits. For data presented in brackets, values are imputed by taking the average nominal value from other observations in the same bracket. Consumer Price Indices collected by Badan Pusat Statistik (the Department of Statistics of Indonesia) for a subset of 82 cities in 2014 are used to deflate nominal income values (BPS, 2021). 2,534 observations (58.01% of the final sample) are unable to be mapped to one of these cities. So, their CPI value is imputed by taking the province's average CPI. This approach is particularly flawed due to the fact that rural observations are more likely to have imputed CPIs that may not accurately capture living expenses in that rural area. However, there exists no other reliable source of price deflators that can be used to adjust for real income.

The main independent variable in our analysis is *parent migrate* which captures whether or not an individual has ever experienced parental migration when that individual was between 5 to 18 years of age. To create the parental migration variable, I link parents' migration histories to the household level to see if any of their children were in the 5 to 18 age bracket during their migration. Children with migratory parents were further categorized into those who were brought along with the parents and those who were left behind by the parent. Parental migration was also further divided into non labor migration and labor migration. Note that these categories are not mutually exclusive as a single child may experience parental migration multiple times with varying circumstances.

3.2 Sample Restrictions

Description	Cumulative Amount	Amount Subtracted
Initial 1993 sample	33,081	
Age 5 to 18	10,952	22,129
Child of HH head	8,997	1,955
Parents alive	8,462	535
Found in 2014	5,860	2,602
Interviewed in 2014	5,586	274
Is employed	4,439	1,147
Has income info	4,368	71

Table 1 – Sample restrictions

Initially, there are 33,081 observations in the first wave of the IFLS fielded in 1993. However, for this study's purposes, I have chosen to restrict my sample in several key ways. First, I only include children aged 5 to 18 in 1993 (33,081 reduced to 10,952). This is meant to capture the most formative years of a child when they are sure to feel the effects of a parent migrating. Second, I only keep individuals that are biological children of a household head (10,952 to 8,997). Third, I exclude children with one or more deceased parents from the sample (8,997 to 8,462). This ensures that comparisons are made across individuals with a similar familial structure. Finally, I link these individuals to their income data recorded in 2014 (8,462 to 5,586) and exclude those who are not working or missing income information (5,586 to 4,368). This leaves us with a final sample size of 4,368 individuals.

To account for the sample selection of the IFLS and attrition of my sample between 1993 and 2014, I will be using the IFLS' longitudinal analysis person weights (PWT14La) to weight my data. These weights capture the inverse probability of being sampled by the first wave of the IFLS and the predicted probability of attrition between the 1993 and 2014 waves imputed using a logit model (RAND, 2020).

3.3 Methods and Model

$$log\ real_i = \beta_0 + \beta_1 parent\ migrate_i + \beta_2 X_i + \mu_i$$

To establish baseline estimates, I run a simple OLS regression with *log real* on the left hand side and *parent migrate* on the right hand side. X contains the following control variables: gender, age in 1993, number of household members in 1993, number of health facilities and schools in the community, and parental education. μ is the error term. Results for this regression are shown in section 4.2 of this paper. However, as mentioned in the literature review, this method suffers from bias due to the endogenous nature of migration.

In the context of this study, unobservable variables may be correlated with both parental migration and future income. In general, these unobservable variables are more likely to bias the coefficient on *parent migrate* upwards. For example, it is possible that parents who are more likely to migrate have a lot of personal connections, even before migrating. These connections can translate to better, higher-paying jobs for their children. Therefore, individuals who experience parental migration are also more likely to earn a higher income regardless of whether their parents' migrate. This implies that the coefficient on *parent migrate* in the simple OLS regression is likely to overestimate the positive effects of parental migration and can be interpreted as an upper bound.

This study seeks to combat the issue of endogeneity by using an Instrumental Variable (IV). Specifically, using IFLS data, I construct the out migration rate of Kabupaten k ($left\ birth_k$) by dividing the number of children who left Kabupaten k by age 12 ($child\ left_k$) by the number of children who were born in Kabupaten k ($child\ born_k$). Then, out migration rates are mapped according to each observation's birth place Kabupaten. A region's out migration rate should be positively correlated with the proportion of children who experience parental migration as a child. The more out migrants there are, the more likely a child in the area experiences parental migration.

$$left\ birth_k = \frac{child\ left_k}{child\ born_k}$$

The two assumptions needed for an IV approach are: (1) that the IV strongly predicts the endogenous variable and (2) that the IV does not directly affect the outcome variable. This first assumption is proven by the first stage regression results available in section 4.3. However, there may be a risk of out migration rates being correlated with future income. This may break the second assumption needed for the IV approach. Specifically, if Kabupatens with high out migration rates have worse living conditions than Kabupatens with lower out migration rates, then out migration rates may be negatively correlated with future income. For example, a Kabupaten with a high out migration rate may have a very limited number of schools and/or healthcare facilities that cause people to leave. This means children who grow up in Kabupatens like this will receive lower inputs in education and health, implying a lower future income. To accommodate this limitation of the IV, I have included community-level controls for the facilities available during childhood.

I have added *num school* (the number of schools in the community per 100 people) and *num health facilities* (the number of health facilities in the community per 100 people) as controls. This data was obtained by linking individuals to their enumeration area that they inhabited closest to age 12. If the individual was not found in the IFLS wave closest to when they were 12, the second closest IFLS wave's value was used. These variables are meant to capture the approximate quality of an individual's childhood community, i.e. the community that they grew up in.

I use the community characteristic specified above instead of an individual's community characteristics from their 2014 location because facility controls from an individual's adult community might not reflect their childhood circumstances. Individuals in our sample may choose to migrate themselves, regardless of parental migration. Since we are concerned that the out migration rate for an individual's birth Kabupaten is correlated to their future income due to childhood living conditions, community characteristics from 2014 will not be a viable way to tackle this issue.

As an alternate approach for an IV, it would be simpler to map each region's recent in migration rates with the child's living location when they were 12 years old and use that as an IV. However, this IV is very likely to be correlated with an individual's future income. A high recent in migration rate may reflect a positive shock in the local economy. Therefore, individuals who grew up in areas with high in migration rates are more likely to earn a higher future income regardless of their parent's migration status. Since there is no straightforward way of controlling for a region's labor market, I opt to use an individual's birth Kabupaten's out migration rate as the IV.

After specifying the IV, we run a two stage least squares regression specified in the equation below. First and second stage results are shown in section 4.3.

$$log\ real_i = \beta_0 + \beta_1 \widehat{parent\ migrate}_i + \beta_2 X_i + \mu_i$$

$$\widehat{parent\ migrate}_i = \alpha_0 + \alpha_1 left\ birth_i + \alpha_2 X_i + \epsilon_i$$

4 Results

4.1 Descriptive Statistics

Sub-group	Entire Sample (n=4368)		Children with Migratory Parents (n=1082)		Children with No Migratory Parents (n=3286)	
Variable	**Mean**	**Std Dev**	**Mean**	**Std Dev**	**Mean**	**Std Dev**
Control Variables						
age_93	11.275	3.742	11.195	3.687	11.299	3.758
male_93	0.596	0.491	0.620	0.486	0.589	0.492
hhldsize_93	5.966	1.977	5.810	1.821	6.012	2.017
num_health_facilities	0.637	0.669	0.600	0.636	0.648	0.678
num_school	0.440	0.463	0.423	0.439	0.445	0.469
mother_educ						
Below Primary	0.176	0.381	0.103	0.305	0.197	0.398
Some Primary	0.627	0.484	0.553	0.497	0.649	0.477
Some Jr High	0.099	0.298	0.148	0.355	0.084	0.278
Some Sr High	0.071	0.257	0.149	0.356	0.048	0.215
Some College	0.022	0.146	0.043	0.204	0.016	0.124
Unknown	0.005	0.073	0.004	0.066	0.006	0.076
father_educ						
Below Primary	0.108	0.310	0.052	0.222	0.124	0.330
Some Primary	0.591	0.492	0.475	0.500	0.625	0.484
Some Jr High	0.111	0.314	0.157	0.364	0.097	0.296
Some Sr High	0.115	0.319	0.182	0.386	0.095	0.294
Some College	0.044	0.205	0.101	0.301	0.028	0.165
Unknown	0.031	0.174	0.034	0.180	0.030	0.172
Migration Variables						
parent_migrated	0.224	0.417	1.000	-	-	-
left_behind	0.093	0.290	0.414	0.493	-	-
self_migrated	0.171	0.377	0.762	0.426	-	-
labor_migration	0.081	0.273	0.362	0.481	-	-
non_labor_migration	0.213	0.409	0.948	0.221	-	-
left_birth	0.058	0.104	0.089	0.168	0.050	0.073
Children's Outcome						
log_real	16.292	1.366	16.495	1.416	16.234	1.346

Table 2 – Descriptive Statistics

Table 2 shows descriptive statistics for the entire sample, children with at least one parent who migrated when the child was between 5 and 18 years old (children with migratory parents), and children without parents who migrated when the child was between 5 and 18 years old (children with no migratory parents).

Compared to individuals who have never experienced parental migration, on average, individuals who have experienced parental migration have a very similar mean age (11.195 years, 11.299 years); higher portion of males (0.620 male, 0.589 male), slightly smaller number of people in a household (5.810 people, 6.012 people), and a lower number of schools (0.423 schools per 100 people, 0.445 schools per 100 people) and health facilities available in their childhood community (0.600 health facilities per 100 people, 0.648 health facilities per 100 people).

Notice that parents of children who experience parental migration between the ages of 5 to 18 generally attain higher education compared to parents of children who did not experience parental migration. For both mothers and fathers, a larger proportion of migratory parents complete Junior High, Senior High, and College compared to non-migratory parents. It is also important to note that the proportion of parents with unknown education are similar for both subsamples. This implies that migratory parents are not more likely to have missing educational information.

In our sample, 22.4% of the observations experience at least one instance of parental migration. Examining the circumstances of parental migration reveal that most of it is driven by non-labor migration (94.8%) and migration where the child is brought along with the parent(s) (76.2%). These categories dominate parental migration compared to labor migration (36.2%) and migration where the child is left behind (41.4%). This contrasts the existing literature that states that labor migration tends to dominate migratory decisions in developing countries. One way to explain this finding is if labor migration is largely made up of young, single, childless migrants. Our sample of older, married parents may not migrate for the same reasons that younger individuals migrate for. Furthermore, this reveals a flaw in the existing literature. Current literature mainly focuses on parental migration where children are left behind and their parents are engaged in labor migration without addressing other contexts of parental migration. Further analysis is needed to categorize the different circumstances of non-labor parental migration.

For the main outcome variable, log of real annual income, we can observe that the average for individuals with migratory parents is slightly higher compared to individuals with non-migratory parents (16.495, 16.234). However, further analysis is needed to attribute this difference in means to the fact that the individual has experienced parental migration. To this end, we run an OLS regression with *log real* as the dependent variable and *parent migrate* as the main independent variable.

4.2 OLS Results

VARIABLES	(1) Simple OLS	(2) OLS w/ parent's educ	(3) OLS w/ more controls
parent_migrated	0.262***	0.0495	0.0256
	(0.0627)	(0.0595)	(0.0587)
father_educ			
Primary School		-0.00968	0.00353
		(0.106)	(0.107)
Jr High		0.259**	0.245*
		(0.129)	(0.130)
Sr High		0.416***	0.424***
		(0.125)	(0.122)
Post-Secondary		0.691***	0.652***
		(0.164)	(0.158)
Unknown		0.0919	0.110
		(0.193)	(0.188)
mother_educ			
Primary School		0.166**	0.179**
		(0.0677)	(0.0703)
Jr High		0.497***	0.517***
		(0.100)	(0.103)
Sr High		0.669***	0.720***
		(0.125)	(0.125)
Post-Secondary		1.009***	1.080***
		(0.159)	(0.167)
Unknown		0.167	0.140
		(0.199)	(0.211)
hhldsize			-0.0153
			(0.0135)
age_93			0.0129**
			(0.00622)
male_93			0.596***
			(0.0526)
num_health_facilities			0.0325
			(0.0832)
num_school			-0.122
			(0.128)
Constant	16.23***	15.95***	15.56***
	(0.0372)	(0.0963)	(0.160)
Observations	4,368	4,368	4,368
R-squared	0.006	0.075	0.123

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3 – OLS Regression (log real income as dependent variable)

Table 3 shows results obtained from a simple OLS regression. Column 1 shows the results of regressing *log real* on *parent migrate* without any controls. Column 2 adds controls for parental education and column 3 further adds household size, age in 1993, birth-assigned gender, number of schools per 100 people, and number of health facilities per 100 people.

The results from column 1 imply that the difference in means for the log of real annual income between the two groups is statistically significant. However, adding controls for parental education reveals that much of this difference can be attributed to the higher education attained by migratory parents. Unsurprisingly, on average, children whose parents are better educated earn a higher real annual income in the future. Adding additional controls for age, gender, household size, and community facilities take away more explanatory power from *parent migrate* indicated by its coefficient becoming closer to zero.

However, this simple OLS approach suffers from bias due to the fact that parental migration is endogenous. As mentioned in the methodology section of the paper, we expect that the coefficient on *parent migrate* in the simple OLS regression is biased upwards. However, interpreting column 3's coefficient as an upper bound may be unwise due to its relatively large standard error. It has a point estimate of 0.0256, a standard error of 0.0587, and a 95% confidence interval of (-0.0899 to 0.1412). Next, we run a two stage least squares regression using *left birth* as an instrumental variable to tackle the problem of endogeneity.

4.3 IV Results

VARIABLES	(1) First Stage
left_birth	0.555***
	(0.0747)
father_educ	
Primary School	0.0491**
	(0.0221)
Jr High	0.119***
	(0.0351)
Sr High	0.152***
	(0.0391)
Post-Secondary	0.248***
	(0.0581)
Unknown	0.0711
	(0.0499)
mother_educ	
Primary School	0.0314
	(0.0204)
Jr High	0.105***
	(0.0400)
Sr High	0.191***
	(0.0448)
Post-Secondary	0.154**
	(0.0755)
Unknown	0.000684
	(0.0888)
hhldsize	-0.00889**
	(0.00416)
age_93	0.00143
	(0.00193)
male_93	0.0213
	(0.0136)
num_health_facilities	-0.0345
	(0.0226)
num_school	0.0261
	(0.0368)
Constant	0.107**
	(0.0421)
Observations	4,368
R-squared	0.083

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4 – First Stage IV Regression (parent migrate as dependent variable)

From the first stage regression results, we can see that the constructed instrumental variable *left birth* is able to accurately predict whether or not an individual has ever experienced parental migration as a child. Holding the control variables constant, on average, an increase of 10 percentage points in an individual's birthplace Kabupaten's out migration rate is associated with a 5.55 percentage point increase in the probability of their parent migrating when the individual was in the 5 to 18 age bracket (95% confidence interval of 4.08 to 7.02 percentage points). Furthermore, performing an F-test on this coefficient yields a value of 55.21. These results confirm that using the Kabupaten out migration rate as an instrumental variable fulfills the first assumption of being able to accurately predict the endogenous variable's variation.

Consistent with the literature on Indonesian migration, individuals with better educated parents are more likely to experience parental migration as a child. Furthermore, the size of a household is negatively correlated with the likelihood of an individual experiencing parental migration.

VARIABLES	(1) Overall Sample	(2) Males Only	(3) Females Only
parent_migrated	0.136	0.606	-0.332
	(0.413)	(0.463)	(0.528)
father_educ			
Primary School	-0.00219	-0.0150	-0.0168
	(0.110)	(0.112)	(0.182)
Jr High	0.230	0.0600	0.475**
	(0.148)	(0.167)	(0.219)
Sr High	0.406***	0.344*	0.421**
	(0.150)	(0.182)	(0.212)
Post-Secondary	0.624***	0.460*	0.775***
	(0.204)	(0.252)	(0.299)
Unknown	0.101	0.127	-0.0349
	(0.199)	(0.217)	(0.323)
mother_educ			
Primary School	0.175**	0.174*	0.191
	(0.0712)	(0.0896)	(0.118)
Jr High	0.505***	0.481***	0.533***
	(0.107)	(0.131)	(0.158)
Sr High	0.698***	0.609***	0.869***
	(0.141)	(0.151)	(0.230)
Post-Secondary	1.063***	0.776***	1.428***
	(0.171)	(0.234)	(0.250)
Unknown	0.139	-0.0270	0.337
	(0.213)	(0.267)	(0.309)
hhldsize	-0.0144	-0.00911	-0.0224
	(0.0136)	(0.0154)	(0.0242)
age_93	0.0129**	0.0134	0.0142
	(0.00622)	(0.00823)	(0.00986)
male_93	0.594***		
	(0.0525)		
num_health_facilities	0.0372	-0.0677	0.280*
	(0.0844)	(0.0956)	(0.161)
num_school	-0.126	0.0178	-0.435**
	(0.130)	(0.151)	(0.209)
Constant	15.55***	16.05***	15.61***
	(0.169)	(0.187)	(0.263)
Observations	4,368	2,569	1,799
R-squared	0.122	0.045	0.083

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5 – Second stage IV Regression (log of real income as dependent variable)

Table 5 shows the second stage regression results using the instrumental variable approach. Column 1 depicts the regression for the entire sample, while column 2 and column 3 show the regression for males and females respectively. This is done to determine if the effects of parental migration differ based on the gender of the child.

We can observe that, for all three regression models, the coefficient on *parent migrate* is statistically insignificant at any relevant confidence level. The standard errors on the *parent migrate* coefficients are extremely large and have wide 95% confidence intervals: (-0.676 to 0.949) for the entire sample, (-0.305 to 1.518) for males, and (-1.370 to 0.707) for females. The differing coefficients on parent migrate may suggest that the effects of parental migration on future income differs based on the child's gender. However, the extremely wide standard errors prevent us from making any conclusive interpretations. Adding province fixed effects and urban/rural dummy variables to the IV regression do not seriously change the magnitude or statistical significance of any of the coefficients (see the appendix for results).

Consistent with the previous section's OLS regression results, higher parental education is associated with higher future income for all three samples, with the effect being stronger for mother's education. Furthermore, age is positively correlated with income and males earn more than females.

4.4 Discussion

Consistent with existing literature, the out migration rate at the Kabupaten level proves to be a good predictor of parental migration, a critical feature of any instrumental variable. The first stage regression results imply that an increase of 10 percentage points in a Kabupaten's out migration rate increases the probability of experiencing parental migration as a child by 5.55 percentage points. However, using a regional out migration rate as an IV may only be viable if the outcome variable is not directly affected by the out migration rate or controls for confounding variables are readily available.

The results from the second stage IV regression show that, for all three samples, the coefficients on *parent migrate* are statistically insignificant at any relevant confidence level. There are a couple of possible explanations for this result.

First, the low R-squared values indicate that a lot of the variation in log of real income remains unexplained by the right hand side variables. This suggests that there exist one or more important omitted variables that can help explain the variation in log of real income. One such variable could be an individual's parental income or household expenditure. Unfortunately, reliable deflators for this data for wave 1 of the IFLS were not readily available. Furthermore, the usage of out migration rates as an IV prevents us from using fixed-effects at a scale smaller than the Kabupaten level. However, a version of this regression is available in the appendix where province fixed-effects and urban/rural dummy variables are included. Unfortunately, none of the coefficients of interest change considerably.

Second, income data often suffer from measurement error and noise; especially when collected via surveys. This widens the standard error on the coefficient on parental migration. For future studies on parental migration, I would recommend examining dependent variables that suffer from less noise, such as educational attainment, measures of health, and subjective wellbeing. Addendum A of this paper shows results from running the same IV regression using a formal sector dummy variable as the main dependent variable (whether or not an individual is working in the formal sector). Addendum B tabulates results from running the IV regression on dummy variables for graduating high school and graduating college.

The results from column 3 of the simple OLS regression and column 1 of the second stage IV regression indicate that older individuals earn more and that males earn a higher income compared to females. Income increasing as an individual gets older could be attributed to some form of tenure or experience at their job. For example, a farmer who has been working for a longer time may have accumulated more land compared to one who is just starting out. Males earning a higher income compared to females could be attributed to the type of employment that males and females participate in. In Indonesian culture, it is often the case that males are expected to be the main sources of income for families, while females are relegated to domestic tasks. It is possible that, when females do work, they are more likely to be engaged in casual, part-time employment, while males are more likely to be a part of more intensive, full-time employment. Here, we assume that casual, part-time employment earns a lower income compared to intensive, full-time employment. For example, a person working part-time by selling leftover nasi goreng is likely to earn a lower annual income compared to someone who dedicates their entire day to working at a nasi goreng cart.

In both regression models, the coefficients on *num school* and *num health facilities* are statistically insignificant at the 10% confidence level, except for the females only IV regression. Surprisingly, the females-only regression finds that income is negatively correlated with the number of schools in the childhood community. We would expect that future income increases if the number of schools available as a child increases. This contradiction suggests that the number of schools available in a community may not accurately represent the quality of the school that the individual had access to. It may be the case that poor individuals in urban areas have a higher number of schools in the area but are restricted to a small subset of lower quality schools due to financial reasons, hence the negative sign on the *num school* coefficient. These results suggest that the number of schools and health facilities may not be a good proxy for the quality of an individual's childhood community. This implies that the correlation between the IV (*left behind*) and the outcome variable (*log real*) may not be resolved by these community-level control variables (*num school* and *num health facilities*). Aside from the statistical insignificance of the main coefficient of interest, this becomes an important caveat to consider when interpreting results from the IV regression.

4.5 Limitations

Due to sample size limitations and restrictions due to the IV, we were not able to run additional analyses to identify the effects of different types of parental migration on a child's future income. Because of this, the results of this study are only meant to measure the aggregate effects of parental migration on an individual's future income. It is important to note that the direction and magnitude of this effect could be affected by the composition of the different circumstances of parental migration. For example, it is possible that, on average, labor migration is associated with better future income due to remittances received as a child. However, non-labor migration generally implies a worse future income due to the absence of a parent without the positive effects of additional income. If we believe this to be the case, then the results of this study are likely to suggest that parental migration has a negative effect on future income due to 94.8% of individuals with migratory parents being subject to non-labor parental migration. If we trust that the IFLS dataset captures an unbiased estimate of the distribution of parental migration circumstances, then the results of this paper may be a good estimate of the aggregate effects of parental migration. But, the results of this study are invalid if the IFLS sample does not capture the true Indonesian population's parental migration behavior.

As previously mentioned in the methodology section, the city-level urban CPIs used to adjust nominal income are far from perfect. Compared to their urban peers, rural observations in our sample are more likely to receive an imputed CPI value that overstates their costs of living. Therefore, due to the inflated CPIs, real income values for these rural observations are likely to be lower than their actual values. This issue can introduce bias to the main coefficient of interest if we believe that parental migration is correlated with the urban or rural status of an individual's location in 2014, which is likely to be the case. The direction of this bias depends on the specific flows of migration that these individuals partake in. However, using these imprecise CPIs is better than comparing raw nominal income data across individuals and localities.

Finally, the IV approach may be flawed due to the presence of correlation between the IV (*left behind*) and the outcome variable of interest (*log real*). We were concerned that Kabupatens with high out migration rates may actually capture the bad quality of facilities at a location. Therefore, out migration rates of birth Kabupatens may be negatively correlated with future income. This correlation becomes an issue if the number of schools and health facilities added as control variables are not able to accurately capture the quality of an individual's childhood community. The messy coefficients on *num school* and *num health facilities* suggest that this is indeed the case.

5 Conclusion

Using the Indonesia Family Life Survey (IFLS) and the out migration rates of a person's birthplace Kabupaten as an instrumental variable, this paper finds inconclusive evidence on the causal long-term effects of childhood parental migration on an individual's log of real annual income as an adult. This can be attributed to the large standard errors that plague the main coefficient of interest. The low R-squared values indicate that the regression models are missing key control variables that can help explain more of the variation in log of real income. It is also likely that income data from surveys suffer from some degree of measurement error and noise.

Consistent with the existing literature, the out migration rate at the Kabupaten level proves to be a good predictor of parental migration. The higher the out migration rate, the more likely an individual experiences parental migration. Also consistent with existing literature, individuals with better educated parents and fewer household members in their childhood are more likely to experience parental migration.

In my final sample, 22.4% of individuals experienced parental migration as a child (1,082 of 4,368). From that subsample, 41.4% had been left behind during the migration, 76.2% had been brought along during migration, 36.2% had been driven by labor reasons, and 94.8% had been driven by non-labor reasons. Contrary to the existing literature, the majority of parental migration experiences in my sample are driven by non-labor reasons. Further investigation is needed to identify specific reasons behind this non-labor migration.

Although the results of this study are inconclusive, they are important to show that extensive longitudinal datasets like the IFLS can be used to answer important questions surrounding the long-term effects of familial migration. Future researchers interested in pursuing this topic are encouraged to examine other outcome variables such as educational attainment, measures of health, and subjective wellbeing.

6 References

1. Antman, Francisca. (2012a). The Impact of Migration on Family Left Behind. *IZA Discussion Paper No. 6374*. http://ftp.iza.org/dp6374.pdf
2. Antman, Francisca. (2012b). Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics*, 25, 1187-1214. https://doi.org/10.1007/s00148-012-0423-y
3. Bai et al. (2017). Effect of Parental Migration on the Academic Performance of Left Behind Children in North Western China. *The Journal of Development Studies*, 54(7), 1154-1170. https://doi.org/10.1080/00220388.2017.1333108
4. Brücker, Herbert and Jahn, Elke J. (2011). Migration and Wage-setting: Reassessing the Labor Market Effects of Migration. *The Scandinavian Journal of Economics*, 113(2), 286-317. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9442.2010.01634.x
5. Cortes, Patricia. (2015). The Feminization of International Migration and its Effects on the Children Left Behind: Evidence from the Philippines. *World Development*, 65, 62-78. https://doi.org/10.1016/j.worlddev.2013.10.021
6. Demurger, Sylvie. (2015). Migration and Families Left Behind. *IZA: World of Labor*. https://wol.iza.org/articles/migration-and-families-left-behind/long
7. Frankenberg, E. and L. Karoly. (1995). The 1993 Indonesian Family Life Survey: Overview and Field Report. RAND, Santa Monica, CA.
8. Hanson, Gordon H. and Woodruff, Christopher. (2003). Emigration and Educational Attainment in Mexico. https://www.researchgate.net/publication/228802899_Emigration_and_Educational_Attainment_in_Mexico

9. Mao, Mingzhi, Zang, Lijun, and Zhang, Haifeng. (2020). The Effects of Parental Absence on Children Development: Evidence from Left-Behind Children in China. *International Journal of Environmental Research and Public Health*, 17(18), 6770. https://dx.doi.org/10.3390%2Fijerph17186770

10. Mckenzie, David and Rapoport, Hillel. (2013). Can Migration Reduce Educational Attainment? Evidence From Mexico. *Policy Research Working Papers*. https://doi.org/10.1596/1813-9450-3952

11. Meng, Xin and Yamauchi, Chikako. (2017). Children of Migrants: The Cumulative Impact of Parental Migration on Children's Education and Health Outcomes in China. *Demography*, 54, 1677-1714. https://doi.org/10.1007/s13524-017-0613-z

12. Nathan, Max. (2011). The long term impacts of migration in British cities: diversity, wages, employment and prices. *SERC Discussion Papers (SERCDP0067)*. http://eprints.lse.ac.uk/33577/

13. Nguyen, Cuong Viet. (2016). Does parental migration really benefit left-behind children? Comparative evidence from Ethiopia, India, Peru and Vietnam. *Social Science and Medicine*, 153, 230-239. https://doi.org/10.1016/j.socscimed.2016.02.021

14. Pardede, Elda Luciana, McCann, Phillip, and Venhorst, Viktor A. (2020). Internal migration in Indonesia: new insights from longitudinal data. *Asian Population Studies*, 16(3), 287-309. https://doi.org/10.1080/17441730.2020.1774139

15. RAND Corporation. (2020). The Indonesia Family Life Survey (IFLS). https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html

16. Strauss, J., F. Witoelar, and B. Sikoki. (2016). The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report. WR-1143/1-NIA/NICHD.

17. Social Science Research Council. (2009). Topic 2 – Remittances and Migration: Conceptual and Methodological Challenges. https://www.ssrc.org/publications/view/EAAEF288-A74F-DE11-AFAC-001CC477EC70/

18. UNESCO Bangkok. (2018). Overview of Internal Migration in Indonesia.

https://bangkok.unesco.org/content/policy-briefs-internal-migration-southeast-asia

19. Wong, Daniel Fu Keung and Chang, Ying-Li. (2010). Mental Health of Chinese Migrant Workers in Factories in Shenzhen, China: Effects of Migration Stress and Social Competence. *Social Work in Mental Health*, 8(4), 305-318. https://www.tandfonline.com/doi/full/10.1080/15332980903217768

20. Yang, Dean. (2008). International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks. *The Economic Journal*, 118(528), 591-630. https://doi.org/10.1111/j.1468-0297.2008.02134.x

21. Zhao et al. (2014). The impact of parental migration on children's school performance in rural China. *China Economic Review*, 31, 43-54. https://doi.org/10.1016/j.chieco.2014.07.013

22. Zhao et al. (2018). Impact of parental migration on psychosocial well-being of children left behind: a qualitative study in rural China. *International Journal for Equity in Health*, 17(80). *https://doi.org/10.1186/s12939-018-0795-z*

23. Zhou, Minhui, Murphy, Rachel, and Tao, Ran. (2014). Effects of Parents' Migration on the Education of Children Left Behind in Rural China. *Population and Development Review*, 40(2), 273-292. https://www.jstor.org/stable/24027923

7 Appendix: IV Regression with province FEs and urban dummies

VARIABLES	(1) First Stage
left_birth	0.476***
	(0.0703)
father_educ	
Primary School	0.0422**
	(0.0211)
Jr High	0.100***
	(0.0343)
Sr High	0.132***
	(0.0377)
Post-Secondary	0.208***
	(0.0547)
Unknown	0.0627
	(0.0492)
mother_educ	
Primary School	0.0301
	(0.0211)
Jr High	0.0917**
	(0.0396)
Sr High	0.183***
	(0.0462)
Post-Secondary	0.165**
	(0.0761)
Unknown	-0.00969
	(0.0855)
hhldsize	-0.0129***
	(0.00421)
age_93	0.00122
	(0.00205)
male_93	0.0156
	(0.0134)
num_health_facilities	-0.0469**
	(0.0200)
num_school	0.0333
	(0.0332)
child_loc	
Small Town	0.0638***
	(0.0208)
Big City	0.00710
	(0.0389)
Constant	0.902***
	(0.0829)
Observations	4,344
R-squared	0.109

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6 – First stage IV Regression (parent migrate as dependent variable)

VARIABLES	(1) Overall Sample	(2) Males Only	(3) Females Only
parent migrated	-0.190	0.441	-0.849
	(0.483)	(0.549)	(0.609)
father_educ			
Primary School	0.0364	0.0503	-0.0433
	(0.105)	(0.113)	(0.180)
Jr High	0.269**	0.143	0.436**
	(0.132)	(0.152)	(0.211)
Sr High	0.421***	0.414**	0.325
	(0.146)	(0.183)	(0.216)
Post-Secondary	0.629***	0.518**	0.657**
	(0.190)	(0.239)	(0.291)
Unknown	0.137	0.193	-0.0659
	(0.197)	(0.218)	(0.312)
mother_educ			
Primary School	0.172**	0.153*	0.241*
	(0.0715)	(0.0896)	(0.126)
Jr High	0.479***	0.432***	0.558***
	(0.110)	(0.135)	(0.167)
Sr High	0.674***	0.548***	0.950***
	(0.151)	(0.162)	(0.252)
Post-Secondary	1.082***	0.741***	1.546***
	(0.182)	(0.237)	(0.296)
Unknown	0.100	-0.110	0.437
	(0.211)	(0.253)	(0.310)
hhldsize	-0.0288**	-0.0181	-0.0420
	(0.0140)	(0.0156)	(0.0255)
age_93	0.0128**	0.0127	0.0162
	(0.00608)	(0.00834)	(0.01000)
num_health_facilities	0.0415	-0.0595	0.274*
	(0.0914)	(0.105)	(0.157)
num_school	-0.0836	0.0488	-0.363*
	(0.134)	(0.157)	(0.215)
child_loc			
Small Town	0.194**	0.101	0.296***
	(0.0807)	(0.101)	(0.111)
Big City	0.247**	0.122	0.444**
	(0.0976)	(0.116)	(0.178)
male_93	0.588***		
	(0.0528)		
Constant	16.00***	16.30***	14.95***
	(0.522)	(0.600)	(0.291)
Observations	4,344	2,549	1,795
R-squared	0.137	0.076	0.068
Province FE	Yes	Yes	Yes

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7 – Second Stage IV Regression (log of real income as dependent variable)

As an attempt to improve the R-squared of the model, we run the original IV regression specified in the methodology section of the paper and add province fixed effects and a childhood location (urban/rural) control variable. Province fixed effects are implemented by using the province that an individual inhabited when they were 12 years old (or closest to 12 years old). Similarly, an individual's childhood location (*child loc*) is categorized as a village, small town, or big city and is recorded from when the individual was 12 years old (or closest to 12 years old).

Based on the first stage IV regression, we can see that adding these two controls does not take away explanatory power from the IV, the Kabupaten-level out migration rate. An increase of 10 percentage points in the out migration rate of an individual's birth Kabupaten is associated with an increase of 4.76 percentage points in the probability that they experience parental migration as a child.

Looking at the second stage IV regression, we can see that adding these two controls improve the R-squared of our model, albeit slightly (from 0.122 to 0.137 for the entire sample). However, this does not change the coefficient of *parent migrated* in any significant way. Although the values of the point estimates change, the extremely large standard errors do not. These results indicate that the model is still missing key variables that could help explain the variation in log of real income.

Adding the two control variables reveal that, for working females, living in a more urban area as a child (small town or big city) is associated with a higher income compared to living in a rural area as a child (village). We also find that the number of household members as a child is weakly negatively correlated with future income; each additional member of an individual's 1993 household is associated with a 2.88 percentage point decrease in real income.

8 Addendum A: Parental Migration on Formal Sector Employment

VARIABLES	(1) Overall Sample	(2) Males Only	(3) Females Only
parent_migrated	-0.209	-0.201	-0.216
	(0.130)	(0.205)	(0.169)
father_educ			
Primary School	0.0388	0.0916**	-0.0366
	(0.0397)	(0.0408)	(0.0680)
Jr High	0.136**	0.148**	0.126
	(0.0526)	(0.0586)	(0.0824)
Sr High	0.166***	0.208***	0.101
	(0.0554)	(0.0681)	(0.0810)
Post-Secondary	0.268***	0.275***	0.253***
	(0.0759)	(0.0977)	(0.0965)
Unknown	0.186***	0.199**	0.163*
	(0.0689)	(0.0977)	(0.0925)
mother_educ			
Primary School	0.0899***	0.127***	0.0382
	(0.0286)	(0.0360)	(0.0473)
Jr High	0.168***	0.238***	0.0674
	(0.0460)	(0.0557)	(0.0684)
Sr High	0.266***	0.298***	0.230***
	(0.0499)	(0.0599)	(0.0826)
Post-Secondary	0.226***	0.178*	0.283***
	(0.0759)	(0.104)	(0.0925)
Unknown	0.276**	0.344**	0.185
	(0.120)	(0.154)	(0.175)
hhldsize	-0.00149	0.00527	-0.0120
	(0.00502)	(0.00586)	(0.00763)
age_93	-0.00831***	-0.00651**	-0.0105***
	(0.00248)	(0.00313)	(0.00362)
male_93	0.104***		
	(0.0209)		
num_health_facilities	-0.0660*	-0.0920***	-0.0156
	(0.0350)	(0.0326)	(0.0570)
num_school	0.0614	0.105	-0.0283
	(0.0668)	(0.0664)	(0.0860)
Constant	0.467***	0.437***	0.664***
	(0.0701)	(0.0770)	(0.109)
Observations	4,368	2,569	1,799
R-squared	0.033	0.028	0.040

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8 – Second Stage IV Regression (formal sector as dependent variable)

In this addendum, we run the IV regression with *formal sector* as the dependent variable. *formal sector* is a dummy variable that is set to 1 if the individual is engaged in any work in the formal sector, i.e. identifying themselves as self-employed with paid workers, government employee, or private employee.

For the entire sample, the point estimate of the coefficient on *parent migrate* is -0.209 with a standard error of 0.130 and a 95% confidence interval of -0.464 to 0.046. The 95% confidence interval suggests that parental migration as a child is likely to be strongly negatively correlated or slightly positively correlated with the probability of formal employment in the future.

Furthermore, parental education seems to have a strong effect on an individual's likelihood of entering the formal workforce. Generally, individuals with better educated parents are more likely to be formally employed. Surprisingly, we find that having a parent with an unknown education increases the probability of being in the formal sector. Age has a very small negative effect, with each additional year of age representing a 0.8 percentage point decrease in the probability of having a job in the formal sector. Interestingly, we also find that the number of health facilities is negatively correlated to the probability of having a formal job, with the effect being strongly statistically significant for males only.

The low R-squared of the model (0.033 for the entire sample) suggests that the current control variables are not great at predicting the variation in formal sector employment. This can be one of the reasons that the coefficient on parent migrate is statistically insignificant at any relevant confidence level. Including additional control variables, such as the parent's own formal sector employment, may improve the explanatory power of the model and reduce the standard errors on the coefficients of interest.

9 Addendum B: Parental Migration on Education

VARIABLES	(1) Overall Sample	(2) Males Only	(3) Females Only
parent_migrated	0.0497	0.0474	0.0577
	(0.121)	(0.168)	(0.161)
father_educ			
Primary School	0.128***	0.148***	0.0876*
	(0.0324)	(0.0389)	(0.0490)
Jr High	0.410***	0.407***	0.411***
	(0.0403)	(0.0525)	(0.0612)
Sr High	0.445***	0.476***	0.389***
	(0.0413)	(0.0528)	(0.0649)
Post-Secondary	0.434***	0.438***	0.432***
	(0.0518)	(0.0721)	(0.0690)
Unknown	0.189***	0.190**	0.168*
	(0.0609)	(0.0775)	(0.0929)
mother_educ			
Primary School	0.156***	0.111***	0.224***
	(0.0264)	(0.0321)	(0.0383)
Jr High	0.346***	0.301***	0.415***
	(0.0387)	(0.0471)	(0.0551)
Sr High	0.404***	0.342***	0.488***
	(0.0434)	(0.0525)	(0.0640)
Post-Secondary	0.420***	0.367***	0.496***
	(0.0402)	(0.0503)	(0.0581)
Unknown	0.323***	0.317**	0.334**
	(0.116)	(0.151)	(0.161)
hhldsize	0.00912*	0.0114*	0.00663
	(0.00516)	(0.00600)	(0.00736)
age_93	-0.00786***	-0.00834***	-0.00728**
	(0.00236)	(0.00287)	(0.00329)
male_93	0.00642		
	(0.0177)		
num_health_facilities	-0.0700*	-0.0535	-0.101
	(0.0424)	(0.0384)	(0.0682)
num_school	0.119**	0.0919*	0.177**
	(0.0523)	(0.0513)	(0.0882)
Constant	0.169***	0.192***	0.145*
	(0.0549)	(0.0632)	(0.0739)
Observations	4,368	2,569	1,799
R-squared	0.239	0.219	0.274

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9 – Second Stage IV Regression (high school grad as dependent variable)

VARIABLES	(1) Overall Sample	(2) Males Only	(3) Females Only
parent_migrated	-0.121	-0.125	-0.0839
	(0.111)	(0.151)	(0.120)
father_educ			
Primary School	0.0228	0.0343*	0.00162
	(0.0169)	(0.0202)	(0.0242)
Jr High	0.171***	0.101***	0.288***
	(0.0323)	(0.0372)	(0.0487)
Sr High	0.251***	0.257***	0.235***
	(0.0431)	(0.0536)	(0.0581)
Post-Secondary	0.447***	0.408***	0.511***
	(0.0597)	(0.0725)	(0.0836)
Unknown	0.0486	0.0438	0.0537
	(0.0433)	(0.0508)	(0.0661)
mother_educ			
Primary School	0.0490***	0.0344**	0.0671***
	(0.0142)	(0.0155)	(0.0222)
Jr High	0.219***	0.203***	0.240***
	(0.0357)	(0.0423)	(0.0549)
Sr High	0.422***	0.397***	0.447***
	(0.0421)	(0.0577)	(0.0643)
Post-Secondary	0.586***	0.533***	0.651***
	(0.0519)	(0.0779)	(0.0690)
Unknown	0.0356	0.0701	-0.0186
	(0.0717)	(0.124)	(0.0757)
hhldsize	-0.00120	0.00264	-0.00559
	(0.00334)	(0.00420)	(0.00503)
age_93	-0.00204	-0.00180	-0.00222
	(0.00159)	(0.00193)	(0.00242)
male_93	-0.0467***		
	(0.0128)		
num_health_facilities	-0.000133	-0.00735	0.0250
	(0.0184)	(0.0175)	(0.0407)
num_school	-0.0248	-0.0172	-0.0418
	(0.0246)	(0.0237)	(0.0513)
Constant	0.108***	0.0537	0.104**
	(0.0337)	(0.0387)	(0.0489)
Observations	4,368	2,569	1,799
R-squared	0.280	0.254	0.332

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10 – Second Stage IV Regression (college grad as dependent variable)

In the first part of this addendum, we run the IV regression with *hs grad* as the dependent variable. *hs grad* is a dummy variable that is set to 1 if the individual has graduated from high school. In the second part of this addendum, we run the IV regression with *col grad* as the dependent variable. *col grad* is a dummy variable that is set to 1 if the individual has graduated from any 2 or 4-year college. Note that the youngest individuals in our sample, aged 5 years old in 1993, would have been 26 years old in 2014, when education data was recorded.

Across both models, for the entire sample, the coefficient on *parent migrate* is not statistically significant at any relevant confidence level. A similar story applies for the male-only and female-only regressions. This can be attributed to the large standard errors on the coefficients.

Unsurprisingly, we find that individuals are more likely to graduate from both high school and college if their parents are better educated. The results also indicate that the number of schools available in an individual's community is positively correlated with their probability of graduating from high school. However, the number of schools in a community is not correlated to an individual's likelihood of graduating from college. For the entire sample, each additional school per 100 people in the community increases the probability of a person graduating from high school by 11.9 percentage points. We also find that age is associated with a very small negative effect on the likelihood of graduating high school, but not college. However, this effect is incredibly close to zero. For the entire sample, each additional year of age is associated with a decrease of 0.786 percentage points to the likelihood of graduating high school. Interestingly, males are less likely to graduate from college compared to females in our sample.